Scripting the role of assessor and assessee in peer assessment in a wiki environment: Impact on peer feedback quality and product improvement.

Abstract

This study investigates how an instructional intervention focused on engaging both the assessor and assessee in the peer feedback process can be advantageous for the quality of students' peer feedback and written product in a wiki-based computer-supported collaborative learning environment in the first year of higher education. The main aim was to examine the effect of structuring the role of the assessee and/or assessor by respectively providing them with a peer feedback request and/or content checklist, together with a structured peer feedback template. The present study adopted a 2x2 design, in which four conditions were compared: (1) a control condition, (2) a feedback request condition, (3) a content checklist condition, and (4) a combination (feedback request + content checklist) condition. Every student (N=125) belonged to a group (n=27) of five and had to fulfil three consecutive assignments, each consisting of writing an abstract for a scientific paper in the wiki. The results revealed that the quality of both peer feedback and the final product increased for all conditions over time, but no significant differences were found between the conditions at time 2 and time 3. However, when the role of the assessee is structured to request for particular peer feedback, this appeared to be favourable for the peer feedback scores, but only at the initial stage of performance. Building on this, limitations, practical implications, and directions for future research are presented.

1 Introduction

1.1 The role of peer feedback in students' learning

In collaborative learning, students work together in small groups to achieve common goals, and in which problems will be tackled more efficiently compared to when students would work individually (eg. Slavin, 1995). Although a peers' competence and lack of objectivity are often mentioned in the literature as examples of possible constraints (eg. Kaufman & Schunn, 2011), research assumes that students acquire more in-group than alone (Dochy, Segers, Van den Bossche, & Gijbels, 2003), which turned out to be also beneficial for students' motivation, social skills, and self-efficacy (Johnson & Johnson, 1994). When students correct the work of other group members and provide feedback in small groups, peer assessment (PA) has proven to be beneficial for the assessor (e.g. Topping, 1998), and assessee (e.g. Tsivitanidou, Zacharia, & Hovardas, 2011), both as an assessment tool and a learning tool (e.g. Evans, 2013). Previous literature on peer assessment emphasises that actively involving students in the assessment process boosts not only their understanding of it (eg. Boud & Molloy, 2013), but that it also increases students' engagement in their own learning (De Wever, et al., 2011). Other research claims that producing feedback reviews "engages students in multiple acts of evaluative judgement, both about the work of peers, and, through a reflective process, about their own work; that it involves them in both invoking and applying criteria to explain those judgements; and that it shifts control of feedback processes into students' hands" (Nicol, Thomson, & Breslin, 2014, p. 102).

When learners are engaged in the role of assessor and assessee, particular PA skills are required. As an assessor, learners need to be able to recognize and assess particular criteria, judging the performance of a peer, and eventually provide peer feedback (PFB). Compared to this, assessees traditionally need to "critically review the PFB they have received, decide which changes are necessary in order to improve their work and proceed with making those changes" (for detailed description see, Hovardas, Tsivitanidou, & Zacharia, 2014, p. 135). PA engages students in cognitive activities such as summarising, explaining, providing feedback, identifying mistakes and gaps, which are dissimilar from the expected performance (Van Lehn, et al., 1995). PFB intends to involve learners by providing and receiving opinions, ideas and suggestions for improvement to their peers (Black & William, 1998).

Previous research claims that online assessment is more beneficial compared to face-to-face assessment (Tsai 2009; Tsai and Liang 2009; Yang and Tsai 2010). A particular reason could be that online technology offers learners more freedom in time and space (Tsai, Lin, & Yuan, 2002). When students are involved in an online PA learning environment, previous research found that students were highly confident and strongly intrinsically motivated (Tseng and Tsai, 2010). In this respect, a growing body of research illustrated that peers' writing performance can be enhanced, when learners are involved learners in online PA activities (eg. Cheng, Liang, & Tsai, 2015; Gielen & De Wever, 2015; Tsai & Chuang, 2013; Xiao & Lucking, 2008). Related to this, online PA appears to be more ideally for augmenting the validity and reliability of PA, when it is integrated into online modules instead of in a traditional setting (Tsai & Liang, 2009; Mostert & Snowball, 2013).

As wikis are free and easily accessible, in which peers can work in group by contributing, commenting and editing further on each other's work, other research claims wikis are ideally suited for collaborative learning (Ebner, Kickmeier-Rust, & Holzinger, 2008). Being an example of an effective online learning environment (Ertmer, Newby, Yu, Liu, Tomory, et al., 2011), wikis are often praised as a computer-supported collaborative learning (CSCL) tool to support PA activities and online collaboration (De Wever, et al., 2011). According to Sadeghi and Kardan (2015), CSCL "can provide a more efficacious, more convenient and more flexible collaborative learning experience for both learners and instructors" (2015, p. 437). More particular, implementing wiki-tasks in higher education appeared to be beneficial for students' learning (De Wever, Hämäläinen, Voet, & Gielen, 2015).

Although feedback has proven to be advantageous for both learning and performance (e.g. Nelson & Schunn, 2008), it appears that not all feedback automatically results in performance improvement (Kluger & Denisi, 1996). Earlier research shows that providing feedback was more beneficial for the assessor's future performance than that of assessees who simply received feedback (Kim, 2009), as students do not necessarily know what to do with the assessor's feedback (Sadler, 1998). Hattie and Gan (2011) also discussed the unpredictable effect of feedback. The framework of Hattie and Timperley (2007), who describe feedback as the component clarifying how well the assesse achieved particular criteria and feed forward as the component offering suggestions how the assesse can actually

improve future performance, may be used to explain this: when feed forward is missing, assessees may not know what to do.

Related to this, previous research already indicated that further research on the impact of PFB on students' learning and performance is needed (e.g. Evans, 2013; Hattie and Timperley, 2007). Inspired by this, the current research attempts to determine *how the roles of assessor and assessee in PA practices should be tailored in order to optimize the PFB process in function of students' learning and performance.*

1.2 Scripting the peer feedback process in function of students' learning

Being a complex learning task, PA requires "high-level" cognitive processing (King, 2002). Nevertheless, previous research has revealed that these particular processes do not happen instinctively (Kollar & Fischer, 2010) and that students require instructional support to participate in these high-level collaboration processes (e.g. Weinberger, Stegmann, Fischer, & Mandl, 2007; Cole, 2009). Related to this, previous research underlines the importance of structure and support in order to safeguard effective feedback (Poverjuc, Brook, & Wray, 2012). For this reason, researchers and instructors should investigate what type of support for the assessor and assessee is fundamental to foster high quality feedback (Hovardas, et al., 2014).

In this respect, scripting is suggested as a possible solution (Fischer, et al., 2013). Collaboration scripts provide more details on role assignment and specify activities in small groups in order to stimulate successful collaborative learning activities (e.g. Kollar, Fischer, & Hesse, 2006). Scripting collaboration has been shown to be beneficial for the acquisition of domain-general skills (e.g. Nooroozi et al., 2013). However, other research revealed contradictory results regarding the effectiveness of collaboration scripts for domain-specific learning outcomes (Kollar et al., 2014). Therefore, investigating the effectiveness of different scripting techniques in various situations is an important field of study (Kollar & Fischer, 2010). Research revealed that collaboration scripts can actually obstruct learner's knowledge attainment when they become too strict or simply too flexible (eg. Fischer, et al., 2013). Therefore, it is essential to determine the precise scripting level that learners require (Dillenbourg, Järvelä, & Fischer, 2009).

Until now, a growing body of research has experimented with varying instructional interventions to enhance the effectiveness of PFB, for instance, by: organizing a training to improve PFB (Sluijsmans et al., 2002), working with multiple raters instead of one (Cho &

Schunn, 2007), offering guiding questions to support the assessor while giving PFB (Gielen & De Wever, 2012), providing sentence openers to encourage interaction between students (Baker & Lund, 1997), or by creating a PFB template with a varying structuring degree to provide feedback and feed forward (Gielen & De Wever, 2015). Nevertheless, the majority of all these instructional interventions are habitually focused on the role of the assessor (Gielen, et al., 2010). It is within this frame that this study is particularly interested in *how collaboration scripts can support the role of both assessor and assessee, in function of students' learning*.

1.3 Scripting the role of both assessee and assessor

In the assessment literature, there is not much information on instructional interventions to enhance the effectiveness of PFB, which focus on the role of the *assessee*. Gibbs and Simpson (2004) mentioned two recommendations for so-called two-stage assignments. Firstly, the assessee should formulate a PFB request in order to receive personalised feedback, e.g. based on an 'a priori question form', in which the assessee specifies his PFB request according to several performance criteria. Secondly, the assessee should also have the opportunity to respond to the received PFB in order to close the feedback loop. Previous research pointed out the advantages of using such a 'posteriori reply form' to close the feedback loop, as it inspires students to reflect on the received PFB and on how they applied it for text revision (Boud, 2000).

As this study focuses on both the role of the assessor and assessee, it is interested in how an instructional intervention, by engaging the assessee in the PFB process, can have an impact on the PFB quality. Therefore, we discuss the 'a priori question form' in more detail. Gielen et al. (2010) incorporated an 'a priori question form', often referred to as '*feedback on demand*' or '*feedback request*', together with a peer feedback template, in an assessment process that encouraged the assessor to fulfil the feedback request of the assessee. It appeared that instructional interventions might not have an effect immediately, but they do have in the long term. Related to the feedback request, the Gielen et al. (2010) wrote that:

> *"Such an intervention may enhance both ''individual accountability'' and ''positive interdependence'' (Slavin, 1989), and motivate and guide assessors to provide 'responsive' feedback (Webb, 1991). It may also result in more appropriate feedback (Webb, 1991) and promote 'mindful reception' (Bangert-Drowns et al., 1991), that is, make assessees feel more personally addressed and subsequently more inclined to apply the feedback"* (Gielen, et al., 2010, p. 308).

A few other examples of studies that embedded a PFB request can be found in the literature, showing a varying degree of delegating responsibility to students. An example to engage

students with feedback was asking students "to specify on their assignment, what they would like feedback on, and giving feedback on nothing else" (Gibbs & Simpson, 2004, p. 24). A similar approach was found elsewhere, requiring students to specify their feedback request in more detail, including for example on which specific aspects they desire feedback (Nicol & MacFarlane-Dick, 2006), or in which specific feedback style (Prins, Sluijsmans, & Kirschner, 2006).

As mentioned earlier, instructional interventions that further specify the role of the assessor in the PFB process, are more widely spread in the assessment literature. Previous research gave an overview of some of these examples such as "coaching by providing hints, prompts, and feedback; modelling the use of cognitive strategies by thinking aloud; presenting cue cards, checklists and process worksheets; asking leading questions; and giving part of a solution" (Van Merriënboer, Kirschner, & Kester, 2003, p. 6). In order to support the assessees' learning, the assessor has a dual task when providing PFB, one that requires high-level cognitive processes. First of all, the assessor needs to deeply process the performance of the assessee and secondly, the assessor needs to engage in planning and monitoring on how to construct valuable PFB for the assessee (Kollar & Fischer, 2010). When assessors are involved in the PFB process, previous research recommends the use of guidelines and checklists (Topping, 2009), as they assist learners to set task-specific goals (Butler, 2002). When the assessor is scripted to complete a checklist, in which essential content concerning a peers' performance is gathered, as a preparation in function of formulating valuable PFB, this instructional intervention could be a possible approach to augment the PFB quality eventually. Previous research underlines that the content of PFB messages is vital for the effectiveness of the feedback (e.g. Cho & MacArthur, 2010). For this reason, it is crucial to further explore alternative instructional interventions, which further specify the role of the assessor and assessee in the PFB process, to shed more light on what kind of support could be necessary to safeguard high quality PFB (Hovardas, Tsivitanidou, & Zacharia, 2014).

2 Methodology

2.1 Participants

All participants in the study were first-year bachelor students of an Educational Sciences program (N = 125), enrolled in the course 'Instructional Sciences'. Participants were randomly assigned to groups (n = 27) of about 4 to 5 students to work on a wiki task, in which each student was required to individually write a number of academic abstracts for scientific articles. From a number of interviews, which are not yet analyzed in detail, a subsample of 43 students indicated that the majority (33/43) already had an experience with PA during their primary or secondary education. However, only a minority (3/43) indicated that providing peer feedback was commonly used. One student explained she gained experience during her teacher training for primary education. For some students (10/43), providing feedback appeared to be something new. The interviews revealed that mostly PA was implemented in secondary education during the evaluation of presentations and group products. It appeared that the presentations were evaluated through both oral and written feedback, while judging group product happened mostly through written feedback. In both cases the feedback was rather qualitative approach, in which students were required to specify positive aspect, but also points of improvement. Only some students (9/43) mentioned they were required to qualitatively evaluate their peers with scores. Only a minority (2/43) actually used assessment criteria to provide a peer with scores from 1 till 5 for each criterion, while other participants (7/43) had to justify their evaluation through qualitative feedback. In this study, no tests reading proficiency and computer literacy were administered.

2.2 Research design and conditions

Each group member had to consecutively read three different academic journal articles and write one abstract for each of these three provided articles (i.e. they received the paper, but the abstract was left out). Each student received three other articles, so there were 15 different articles in each group. For each abstract, students participated in four phases: (1) writing a draft version of the abstract (2) providing non-reciprocal PFB to (and also receiving PFB from) another group member (the same one for the three iterations), (3) revising the draft version based on the feedback to construct a final version, and (4) evaluating the received PFB. All four phases were part of their curriculum requirements.

During the PFB process, the instructor provided the same structured PFB template to every student. This template consisted of a list of seven criteria (intention of research, problem statement, methodology, results, conclusion, limitations, and general judgment), which was structured in a way that students were encouraged to provide both PFB and feed forward for each criterion separately (Hattie & Timperley, 2007). During the first PFB cycle, only the assessor was instructed to employ this PFB template, while the assessee was not involved in the PFB process. As part of the instructional intervention, this study deliberately introduced additional structuring during the second and third feedback cycle in order to engage both the assessee and assessor in the PFB process. To examine the effects of this script, this study adopted a quasi-experimental 2 x 2 factorial design, which resulted in four conditions: (1) a control condition, in which only the PFB template was provided; (2) a request condition, in which in addition to the PFB template, the assessee was provided with peer feedback request form, which required the assessee to formulate a specific feedback demand; (3) a content condition, in which again the same PFB template was provided, but this time together with a content checklist form, which required the assessor to actually select the essential content from the paper, meaning that the content checklist was contextualized to a specific paper; and (4) a combination (request + content) condition, in which the same PFB template was combined with the peer feedback request form (i.e. structuring the role of the assessee) and a content checklist form (i.e. structuring the role of the assessor).

To sum up, the PFB process in this study consisted out of 4 phases, which are also illustrated in detail in Figure 1: (1) requesting for specific PFB, (2) preparing PFB content, (3) providing feedback on a structured PFB template, based on (a) the PFB request and/or (b) the prepared content checklist, and finally (4) evaluating the received PFB. In the appendix, screenshots provide a visual overview of all the used templates and different steps of one feedback cycle. In order to study the differential impact of the script that further specifies the role of the assessor and assessee during the different PFB phases, this study primarily focuses on the first two PFB phases, in which (1) the assessee requests for particular PFB and (2) the assessor prepares PFB content with the help of a content checklists. Figure 1 gives an overview of the specific procedure for each condition.

<< Figure 1. Procedure for each condition >>

2.3 The role of the instructor

In the practical part, the instructor facilitates the learning experience by managing the online learning environment, in which he can for example appoint or randomize students into different groups, decide on certain wiki settings such as privacy, create a group forum, make scientific articles or group documents available for students. First of all, the instructor gave an introduction to this task during a face-to-face lecture of 60 minutes at the start of the academic year. Additionally, the practical website gave an overview of the whole assignment, in which students could find additional information such as FAQ, but also an instructional video, in which all the necessary steps of the wiki assignment are chronologically discussed. During the assignment period, the role of the instructor also requires mentoring and monitoring on a class, group and student level. On the group level, the instructor is required for example to mediate and intervene if particular group members do not respect the deadlines and to monitor if all group members actively participate and contribute to the wiki assignment. On a personal level, the instructor needs to deal for example with students who drop out or for who the assignment seems too overwhelming. As these students are first-year bachelor students, the instructor will offer regularly instructions by posting weekly announcements in the online learning environment. Students were encouraged to seek help in first instance with students within their group or between other groups. Finally, students could mail their unresolved practical questions to the instructor.

2.4 Instruments

2.4.1 Peer feedback request

After finishing the draft version of an abstract, students in the request conditions are required to compile a specific PFB request or 'feedback on demand' (Gielen, et al., 2010), based on the expectations and judgment of their own work. In particular, the assessees have to indicate firstly on which of the seven predetermined criteria they expect feedback and secondly, what kind of specific feedback they expect. When clarifying their specific PFB request on the list of selected criteria, the assessees are left freely regarding how they specifically ask for feedback. This could be done for example by formulating specific questions, or by referring to a particular paragraph, etc. As illustrated in Figure 1, the PFB request has to be submitted before the assessor can start compiling and providing his/her PFB regarding the draft version of the assessee.

2.4.2 Content checklist for the assessor

In order to activate the role of the assessor, students in the content checklist condition are required to complete a checklist, based on the same seven predetermined criteria, so as to be able to deeply process a peers' performance (Kollar & Fischer, 2010). In order to do so, students start off by reading the scientific article of the peer, for who they have to provide PFB in the next phase. Afterwards, students are encouraged to carefully identify the information, which according to the assessor contains the essential components of the scientific article that should be present in the abstract, to meet the expectations of good performance. Next, students are required to sum up and categorise the selected content for all criteria separately. The assessor is scripted to complete this content checklist before actually formulating PFB, as its purpose is to serve as an input source during the third phase in the PFB process.

2.4.3 Structured peer feedback template

For the third phase in the PFB process, the instructor provided all four conditions with an identically structured PFB template, consisting out of four sections: (1) a list of predetermined criteria (2) a section to provide feedback on each criterion, (3) a section to provide feed forward on each criterion and finally, (4) a section to evaluate the received PFB for each criterion. Following, these four sections will be discussed in further detail.

The first section provides a list of seven predetermined criteria, presented in a table. In the second and third section, students were required to provide feedback and feed forward: "How am I going?" (What progress is being made toward the goal?), and "Where to next?" (What activities need to be undertaken to make future improvement?) (Hattie & Timperley, 2007, p. 86). While providing PFB on the structured peer feedback template, the assessor needs to check whether the assessee previously submitted a specific peer feedback request, and, if so, provide specific feedback and feed forward exclusively focussing on the requested criteria, instead of providing feedback and feed forward related to all of the criteria. In addition, the assessors of the content checklist conditions had to actively employ the information of the checklist, when formulating their feedback on the structured peer feedback template. In the fourth section, students were instructed to evaluate their received PFB. The assessees could indicate, for each criterion and its corresponding feedback, whether the received PFB was valuable or rather irrelevant.

2.4.4 Scoring rubric for quality of peer feedback messages (Peer feedback score)

The Feedback Quality Index (Prins et al., 2006) was used to measure the quality of peer feedback messages. Building on this, a series of previous studies (Gielen & De Wever, 2012; 2015) and the present study consistently incorporated this scoring rubric, maintaining all three main categories (use of criteria, nature of the feedback, and writing style), and their sub-categories with corresponding scoring percentages (Prins et al., 2006). This resulted in a scoring rubric of 9 items with a scoring range between 0 and 100 to measure the quality of PFB messages of first-year higher education students. (For more info, see Gielen & De Wever, 2015).

2.4.5 Scoring rubric for quality of the wiki product (Product score)

This study applied a scoring rubric, used previously in two studies (Gielen & De Wever, 2012; 2015), in which the necessary components of a good abstract are integrated. This scoring rubric counted four main categories (situating the study, content of the abstract, style, and general impression) and nine corresponding sub categories. (For more info, see Gielen & De Wever, 2015). Similar to the previous rubric, this scoring rubric for analysing the quality of the wiki product also had a scoring range between 0 and 100.

2.4.6 Students' perception towards PA at the end of the assignment

At the end of the wiki-assignment, students were asked to fill out a questionnaire. Items were measured using 5-point Likert scales, from 1 (totally disagree) to 5 (totally agree). The first main section 'The role of the assessor' consisted of 'Perception towards providing feedback' (e.g., "I believe it is disadvantageous that the assessment happens in an online environment") (7 items), 'Perception towards the PFB request' (e.g., "I believe the PFB request limits me as an assessor because habitually I would provide more PFB") (7 items) and 'Perception towards the content checklist' (e.g., "I believe the content checklist is useful to increase the quality of my PFB") (5 items).

The second main section 'The role of the assessee' attempts to shed more light on the perception of students when they act as assessees. The questions dealt with: 'Perception towards receiving feedback' (e.g., "I believe that the value of the received feedback from an assessor or instructor is equal") (5 items), 'Perception towards the PFB request' (e.g., " I believe the PFB request is an added value for the whole process") (7 items) and 'Evaluation

of the received feedback' (e.g., "I believe that generally the quality of the received feedback is insufficiently") (5 items). For all the different items, analyses of variance were performed to compare the effect on students' perception in the four conditions.

2.5 Hypotheses

The following hypotheses are put forward with respect to the PFB quality scores.

Since at time 1 no additional forms are introduced and all conditions are only using the PFB template, (H1) we expect the PFB quality scores of students in all four conditions not to differ significantly at the start of performance. With respect to the quality of the PFB over time, (H2A) we expect the PFB quality scores to increase significantly for all four conditions from time 1 to time 3, as students become more proficient assessors when they are involved in multiple practice occasions (e.g. Van Steendam, Rijlaarsdam, Sercu, & Van den Bergh, 2010). Thus, we also expect a significant increase of PFB scores (H2B) from time 1 to time 2 and (H2C) from time 2 to time 3, for all conditions. As the intervention was introduced during the second iteration, we expect the PFB quality scores of students in the request, content, and combined condition to be higher than those of students in the control condition (H3A) at time 2 and (H3B) at time 3, since previous research has shown that a certain degree of structure can have a beneficial effect on the PA process (Gielen & De Wever, 2015). More specifically, the peer feedback request form is supposed to increase the PFB scores significantly, as it motivates and directs the assessor to formulate more 'responsive' feedback-on-demand (Webb, 1991). As previous research recommended the use of guidelines and checklists when assessors are involved in the PFB process (Topping, 2009), the content checklist is supposed to increase the PFB scores significantly, as assessors are more profoundly prepared and encouraged to formulate high quality PFB content, which is in turn essential for the effectiveness of feedback (e.g. Cho & MacArthur, 2010). With respect to the combined request + content condition, no specific hypothesis is formulated, as there was no literature available that has investigated this combination.

Similar hypotheses are formulated with respect to the product scores. Since at time 1 no additional forms are introduced, and all conditions are only using the PFB template, we also expect (H4) the product scores of students in the four conditions to be the same at the start of performance. With respect to the quality of the writing product over time, we expect (H5A) a

significant increase of the product scores for all four conditions from time 1 to time 3, as practice results in performance improvement over time (eg. Gielen & De Wever, 2015). Thus, we expect that results increase significantly (H5B) from time 1 to time 2 and (H5C) from time 2 to time 3. When students receive more high quality feedback, we assume that this will have a stronger impact on the final writing product, as other research claims that more specific and elaborated feedback stimulates better performance and outcomes (eg. Strijbos, et al., 2010). As the intervention was first introduced in the second iteration, we expect the product scores of students in the request, content, and combined condition to be higher than those of students in the control condition (H3A) at time 2 and (H3B) at time 3. More specifically, the peer feedback request form is supposed to increase the product scores significantly, since previous research has revealed that assessees pay more attention on feedback-on-demand (Gielen, et al., 2010) and response - specific feedback augments learning efficiency (Hansen & Almond, 2007). Following, the content checklist is supposed to increase the product scores significantly, as we believe that high quality PFB content has the potential to improve the writing product more compared to low quality PFB content. However, we need to take into account that all kind of feedback can be beneficial for students' performance (Topping, 1998). With respect to the combined request + content condition, no specific hypotheses is formulated for the combination.

2.6 Data analysis

Given the clear hierarchical structure of the data, namely three measurement occasions (i.e. the PFB moments, indicated by the variable 'time' (level 1) are nested within each of the 125 students (level 2), who are in turn nested within 27 groups (level 3), multilevel modelling for repeated measures (MLwiN 2.29) was used to analyse the PFB quality and the product quality (ie. the quality of the versions of the abstract written in the wiki). Firstly, a fully unconditional null model was tested for both PFB score and product. Following this, a compound symmetry model was created by adding the categorical predictor 'time' to the null model, since it is a random intercept model with no explanatory variables except for the measurement occasions (Snijders & Bosker, 1999). In this model, time 1 is taken as the reference category. Then, the categorical predictor 'condition' is added in the next step for both the PFB and product score. In a final phase, the interaction condition*time was added to the model. By using a stepwise multilevel approach, we could check the additional value of each subset of variables to the model.

3 Results

3.1 Peer feedback score

The null model shows that 24.98% of the total PFB variance is situated at the group level (p=.011), the proportion of variance due to differences between students within groups was 35.81% (p<.001), and finally 39.21% of the total variance is situated at the time level (p<.001; see also Table 1 in appendix). Related to hypothesis 1, results revealed that at the start (i.e. time 1, when the structuring intervention had not yet started), the request condition has significant lower PFB scores compared to the content condition ($\chi2$=7.192, df=1, p=.008) and compared to the combination condition ($\chi2$=6.326, df=1, p=.012), which is in contrast to H1.

<< Figure 2A. Evolution of peer feedback scores over time. Note: full lines indicate significant increases, dashed lines indicated non-significant increases at p=.05. At time 1 there was also a significant difference between the request condition and both the content and combination condition. >>

Figure 2A shows that the PFB scores increase from time 1 to time 3 for all four conditions. When taking a detailed look at the separate conditions, only the scores of the control ($\chi2$=6.385, df=1, p=.011) and request condition ($\chi2$=13.969, df=1, p<.001) increased significantly from time 1 to time 3, which only partly confirms hypothesis 2A. As shown in Figure 2B, the PFB scores increased significantly from time 1 to time 2 for the request condition with 11.88 ($\chi2$=16.381, df=1, p<.001), and with 7.50 for the combination condition ($\chi2$=5.936, df=1, p=.015), which only partly confirms hypothesis 2B. When taking a closer look at the increase in PFB scores from time 2 to time 3, results only show a significant increase for the control condition with 5.94 ($\chi2$=3.969, df=1, p=.046), which only partly confirms hypothesis 2C. When taking a closer look at the PFB scores at time 2 and time 3, no significant differences were found between the four research conditions, not confirming hypothesis 3A and 3B.

<< Figure 2B. Peer feedback scores at time 1, time 2 and time 3. Note: full lines indicate significant increases, dashed lines indicated non-significant increases at p=.05. >>

3.2 Product score

As shown in Table 2 in the appendix, the null model reveals that 1.28% of the total variance of product scores is situated at the group level (p=.684), the proportion of variance due to differences between students within groups was 15.36% (p=.021), and finally 83.36% of the total variance is situated at the time level (p<.001). Looking at hypothesis 1, results revealed no significant differences in product scores between the four conditions at the start (i.e. time 1, when the structuring intervention had not yet started), which confirms hypothesis 4.

<< Figure 3A. Evolution of peer feedback scores over time. Note: full lines indicate significant increases, dashed lines indicated non-significant increases at p=.05. >>

Figure 3A shows that all conditions improve significantly over time. In more detail, the product scores increased significantly for the control ($\chi2$=18.484, df=1, p<.001), request ($\chi2$=7.868, df=1, p=.005), content ($\chi2$=11.667, df=1, p<.001), and combination condition ($\chi2$=6.037, df=1, p=.014) from time 1 to time 3, which confirms hypothesis 5A. When decomposing the increase over time, results revealed differences between the four conditions regarding the increase of product scores over time. As shown in Figure 3B, the product scores of three out of four conditions increased significantly after the structuring intervention, which confirms hypothesis 2B, except for the content condition. More specifically, the conditions for who the product scores increased significantly from time 1 to time 2 were the control condition with 9.06 ($\chi2$=9.572, df=1, p=.002), the request condition with 6.82 ($\chi2$=5.587, df=1, p=.018), and the combination condition with 8.70 ($\chi2$=8.270, df=1, p<.004). The content condition increased with 5.40 from time 1 to time 2, which is close to significant ($\chi2$=3.186, df=1, p=.074). Although three out of four conditions had higher product scores at time 3, no significant increase was found for any condition from time 2 to time 3, which does not confirm hypothesis 5C. When taking a closer look at the product scores at time 2 and time 3, no significant differences were found between the four research conditions, contrary to hypothesis 6A and 6B.

<< Figure 3B. Product scores at time 1, time 2 and time 3. Note: full lines indicate significant increases, dashed lines indicated non-significant increases at p=.05. >>

3.3 Students' perception at the end of the wiki-task

3.3.1 Students' perception towards PA

In order to provide some extra background information, a questionnaire revealed that mostly all students had gained some experience with PA before arriving at university (M= 3.07, SD=1.24). In general, providing feedback was not perceived as a difficult task (M= 2.46, SD=0.92). Also, students not perceive it as a shortcoming when the PA process happened in an online setting (M= 2.38, SD=1.08). When evaluating their received PFB, students indicated that they did not perceive the quality as insufficient (M= 2.24, SD=0.98) and they are more or less convinced that the quality of their final version enhanced significantly, thanks to the received feedback (M= 3.50, SD=0.92).

3.3.2 Students' perception towards the additional structure in PFB process

In general, students more or less agreed that the PFB request was useful for formulating PFB in their role as assessor (M= 3.61, SD=0.97). Further analysis revealed significant differences between the conditions [$F(3, 114) = 3.78$, $p = 0.012$]. More particular, it appeared that both the request (M= 3.90, SD=0.83, p=.043) and content (M= 3.86, SD=0.87, p=.024) condition perceived the PFB request significantly more useful to formulate feedback, compared to students in the combination condition (M= 3.21, SD=1.14). No other significant differences were found compared with the control condition (M= 3.45, SD=0.87). Moreover, when a PFB request is part of the PA process, students believe they would provide more specific feedback in their PFB messages (M= 3.82, SD=0.83). Additionally, it appeared that these messages would comprise more suggestions on how to improve future work (M= 3.63, SD=0.80), revealing significant differences between the conditions [$F(3, 112) = 2.85$, p= 0.040]. Especially students in the content condition (M= 3.83, SD=0.65, p=.043) demonstrated they would provide significantly more suggestive feedback, compared to students in the combination condition (M= 3.28, SD=0.88). In comparison with the control (M= 3.64, SD=0.82) or request condition (M= 3.77, SD=0.77), no other significant differences were found. Similarly, students perceived the content checklist as a helpful instructional intervention to formulate PFB about a peers' draft version (M= 4.05, SD=0.80). In general, students more or less agreed that the content checklist actually helps the assessor to increase the quality of the PFB (M= 3.26, SD=0.95).

4 Discussion

4.1 Peer feedback quality

Taking into account that students had previous experience with PA and they did not perceive offering PFB as a difficult task, the results revealed that the PFB quality scores, measured by the Feedback Quality Index (Prins et al., 2006), increased for all students over time after multiple practice occasions. Also, being engaged in an online setting appeared not to be a problem for students, which supports research that acknowledges the beneficial effect of engaging students in online PA activities (eg. Cheng, Liang, & Tsai, 2015; Gielen & De Wever, 2015; Tsai & Chuang, 2013; Xiao & Lucking, 2008). However, only the control and the request condition increased significantly from time 1 to time 3. This finding suggests that by involving students in similar practice occasions, PFB quality scores could improve significantly over time. This is in line with earlier research (eg. Van Steendam, Rijlaarsdam, Sercu, & Van den Bergh, 2010; Gielen & De Wever, 2015) claiming that practice is perceived as a condition sine qua non, in order to evolve as a skilled peer assessor. While there are studies indicating that students do not really require training in assessment (eg. Cho & MacArthur, 2010), most research stresses that feedback practices in higher education should take into account more practice occasions in providing and receiving of feedback (eg. Nicol, 2010). Although there were unexpected significant differences in PFB scores between the four conditions at the start of the intervention, results reveal that over time these significant differences are eliminated when students have multiple occasions to practice a similar performance.

As earlier research has shown that a certain amount of structure is beneficial for the PA process (eg. Gielen & De Wever, 2012; 2015), we take a closer look at what happens when students become more actively involved at time 2 in the PFB process, by focusing on the one hand on the assessee who formulates a specific peer feedback request for different criteria, and, on the other hand, on the assessor who prepares a content checklist as input source when formulating peer feedback. Completing a PFB request appeared to be a useful approach to increase the PFB quality, as both conditions that incorporated a PFB request in the PFB process had a significant increase in PFB quality scores from time 1 to time 2, with or without the assessor completing the content checklist. When the assessee has the opportunity to ask for specific PFB, the assessor could be more inclined to provide higher quality PFB on topic. More particular, findings showed that students in the content condition believed they would

offer significantly more suggestive feedback, when a PFB request would have been implemented in their PFB process, compared to the students in the combination condition. One reason for this could be that students, who employed a content checklist, have collected a lot of information during their preparation. When responding to a particular PFB request, they believe they would provide more suggestive feedback in function of future improvement. On the other hand, we need to consider that students in the combination condition could perceive the PFB request as a restriction when composing their PFB messages. Related to this, students indicated not to feel restrained when providing feedback on demand, and it became clear that they offered feedback on more aspects, than merely responding to the PFB request. This is in line with research of Gielen et al. (2010), which claims that the implementation of 'feedback on demand' could motivate and direct the assessor to provide more 'responsive' feedback (Webb, 1991), and thus provide PFB of a higher quality, driven by a higher 'individual accountability' and 'positive interdependence' (Slavin, 1981). From time 2 to time 3, only the control condition showed a significant increase while both conditions that incorporated a PFB request demonstrated a minor decline in PFB scores at time 3. This finding suggests that a PFB request could be useful approach to boost PFB quality scores, especially in the initial phase of performance, but that its effect could diminish over time. Related to this, it appeared that students generally acknowledged the usefulness of a PFB request during the assignment period. More particularly, the request and content condition perceived the PFB request significantly more useful compared to the combination condition. No significant differences were found in comparison with the control condition.

Future research should confirm whether the PFB request is actually able to boost the PFB scores, at least in the request condition, or whether this increase from time 1 to time 2 may be due to the significantly lower scores at time 1. Since students were randomly assigned to conditions, this different starting level was not expected and is a limitation of the study in the sense that it hinders a straightforward interpretation of the increase. Future research, replicating these conditions, may be necessary to confirm this hypothesis. When only the assessor is scripted to complete the content checklist before actually providing PFB, without being asked for specific feedback by the assessee, results pointed out no significant increase over time, taking into account that the content condition had generally high PFB quality scores at all times. Yet, results revealed no significant differences in PFB scores between the four conditions neither at time 2 or time 3. Also, given that both conditions that implemented the content checklist started at a rather high level (in spite of the random assignment of

individuals to groups and the random assignment of groups to conditions), this may be the cause of the non-significant increases over time. Therefore, results are rather inconclusive to determine the actual effectiveness of the content checklist for the PFB process. Nevertheless, the questionnaire revealed that assessors perceive the content checklist as a helpful instructional tool, which they believe could have the potential to increase the assessor's PFB quality.

4.2 Product quality

When looking into the product quality scores over time, there was an increase from time 1 to time 3 with respect to the writing products for all students. Similar to the PFB scores, this finding suggests that when students have the opportunity to gain experience in a similar performance at multiple practice occasions, the quality of the product will increase generally as well. This result is in line with a previous study, which claims that practice leads to product improvement over time (Gielen & De Wever, 2015).

When decomposing the increase over time, results showed that the product scores of three out of four conditions increase significantly from time 1 to time 2, when assessee and assessor are more involved in the PFB process regarding a similar assignment. This is in agreement with research, which encourages the engagement of peers, both assessor and assessee in PA (e.g. Evans, 2013), as it requires students to deal with the essential aspects that correspond with high quality performance (Topping, 1998). Interestingly, completing a PFB request appeared to be a useful approach once again, not only to increase the PFB scores in the initial phase of performance, but as well to boost the product quality from time 1 to time 2, whilst both conditions that incorporated a PFB request had a significant increase in product scores, with or/and without the assessor completing the content checklist. This finding suggests that receiving specific feedback on request could have the potential to boost the quality of the peer feedback but also the quality of the actual writing product, since assessees could receive more detailed answers on particular previous issues. Gielen, et al. (2010) showed that the question form, in which secondary school students could request for feedback, led to more effective feedback. A possible explanation could be that "the assessee may in return pay more attention to the feedback that refers to these personal questions and this 'mindful reception' is crucial to the feedback's instructional value" (Gielen, et al., 2010, p. 158). This is in line with research, which suggests providing feedback on request as an approach to encourage students to make

more effective use of their received feedback (Gibbs & Simpson's, 2004). When the assessor can give feedback on request, this could support the assessee to augment the quality of the (writing) product. This is in line with research of Shute, Hansen, and Almond (2007), showing that response - specific feedback appears to augment learning efficiency, while other research revealed that more specific and elaborated feedback stimulates better performance and outcomes (Strijbos, et al., 2010).

Since the product scores of the control condition showed a significant increase from time 1 to time 2 as well, without any additional structure or support, besides the provided structured peer feedback template, this finding may question the actual necessity of providing additional structure in the PFB process. Findings indicated that all students were generally satisfied with the quality of the PFB they received. They believed that their PFB was satisfiable underpinned with specific article content or links to the article. Finally, results revealed that the assessees perceived the received PFB to augment the quality of their final version. This finding is in agreement with research, which advocates that every variety of feedback, whatever its amount or specificity, can have a positive effect on students' product scores (Topping, 1998). We should also take into account that all students, including the control condition, were using the same structured PFB template. In this way, the PFB process was already structured to a certain extent.

From time 2 to time 3, no significant increase was found for any of the four conditions after one more practice occasions. Other research showed that PFB does not necessarily augments performance in a later phase of PA activity (Chen & Tsai, 2009), when students are involved in multiple practice occasions. Only the combination condition demonstrated a minor decline in PFB quality, but we have to take into account that their results were the highest of all conditions at time 2. It may be that after similar practice occasions, the impact of the feedback request is diminishing. It appears that this strategy is especially useful to boost PFB and product scores in the initial phase of practice, but also that the effect diminishes after more practice occasions. This is in line with research, which claims that students' engagement in producing feedback reduces their own need for external feedback (Nicol, Thomson, & Breslin, 2014). For this reason, we believe that adaptable external collaboration scripts could be a possible solution, in which provided structure could be faded in or out over time, according to students' needs (eg. Wecker & Fischer, 2011). Following, we need to mention that the content condition had no significant increase from time 1 to time 2, or from time 2 to

time 3, but this condition showed nevertheless a significant improvement from time 1 to time 3, and had even more the highest product scores of all conditions at the end of performance. Finally, results revealed no significant differences in product scores between the four conditions neither at time 2 or time 3.

5 Conclusion, limitations, and direction for future research

As students perceived being engaged in online PA activities not as a difficult or time-consuming task, this study illustrated that offering additional structure in PA, to further specify the role of the assessee and assessor during the PFB process, is a valuable approach to increase both the quality of PFB and the writing product. Students indicated they were generally satisfied with the quality of the peer feedback, which appeared to be specific and elaborated with relevant information to improve future performance. Over time, results pointed out that all students improved after multiple practice occasions in providing PFB of a higher quality and delivering performances of a significant higher quality. It became clear that the additional structure for both assessee and assessor could be useful, especially in the initial phase of performance, but also that its effect diminishes over time. This study showed that students in their role as assessor especially appreciated the effectiveness of the PFB request and content checklist in the PFB process. Related to this, the assessees indicated to believe that the PFB has the potential to increase the quality of their future performance. For this reason, the data suggests that providing additional structure, such as a content checklist for the assessor and a PFB request for the assessee is a useful approach to actively engage students in writing and assessment activities and boost the quality of both PFB and product at the initial stage of performance.

Situated in an authentic learning environment, we have to take into account that not all contextual factors can be controlled for and that for example students' metacognitive approach, reading proficiency or computer literacy can affect the results. Although we tried to control this as much as possible by dividing students randomly over the different conditions, future research may consider taking into account these factors. As writing an abstract is a specific task, we have to consider that a significant improvement in product quality from draft to final version could be caused by the received PFB, but as well by experience through multiple practice occasions (Kluger & Denisi, 1996). Related to this, detailed analyses of the

specific content of the PFB request and checklist could shed more light on the PFB quality. Based on the recommendations of previous research (Gielen & De Wever, 2015), all students received an identical PFB template with a particular structuring degree for completing the third PFB phase. Thus, all conditions were structured to a certain extent in the PFB process. On one hand, structuring the PFB to a certain extent seems to be promising approach, given the overall increase in product scores over time, but, on the other hand, we are unable to fully determine which part of the product improvement is triggered by the quality of the peer feedback, by the received structure in the peer feedback process, or merely by practice in similar performance.

More research on optimizing the peer feedback process in function of students' learning is necessary. Future studies could replicate this study in order to examine the differential effect when the provided additional structure is kept or faded out at time 3, or when students have simply the option to create their own personalised script, in which they can select the required steps in the PFB process. As an implication for practice, instructors wishing to implement peer assessment could consider the following recommendations. Firstly, this study recommends implementing a structured PFB template (Gielen & De Wever, 2015) to guide students through the PFB process, in which they need to provide feedback and feed forward on a list of pre-specified, or preferably mutual discussed criteria (Sluijsmans, 2002), but they also need to evaluate the feedback in order to close the feedback loop (Boud, 2000), as illustrated in the appendix. Secondly, this study advises to foresee multiple practice occasions or feedback cycles, when instructors are planning to involve students as assessors and assessees in the PFB process.

References

Baker, M. J., & Lund, K. (1996). Flexibly structuring the interaction in a CSCL environment. In P. Brna, A. Paiva & J. A. Self (Eds.), *European conference on artificial intelligence in education* (pp. 401–407). Lisbon, Portugal.

Bangert-Drowns, R., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61,* 213-238.

Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice, 5*, 7–74.

Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society *Studies in Continuing Education, 22*, 151-167.

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: the challenge of design, *Assessment & Evaluation in Higher Education, 38*, 698-712.

Butler, D. L. (2002). Individualizing instruction in self-regulated learning. *Theory and Practice, 41*, 81-92.

Chen, Y. C., & Tsai, C. C. (2009). An educational research course facilitated by online peer assessment. *Innovations in Education and Teaching International, 46*, 105–117.

Cho, K., & MacArthur, C. (2010) Student revision with peer and expert reviewing. *Learning and Instruction, 20*, 328-338.

Cho, K. & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline, *Computers and Education, 48*, 409–426.

Cole, M. (2009). Using Wiki technology to support student engagement: Lessons from the trenches. *Computers and Education, 52*, 141–146.

De Wever, B., Hämäläinen, R., Voet, M., & Gielen, M. (2015). A wiki task for first-year university students: The effect of scripting students' collaboration. *The Internet and Higher Education, 25,* 37-44.

De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2011). Assessing collaboration in a wiki: The reliability of university students' peer assessment. *The Internet and Higher Education, 14,* 201-206.

Dillenbourg, P., Järvelä, S., & Fischer, F. (2009). The evolution of research in computer-supported collaborative learning: From design to orchestration. In N. Balacheff, S. Ludvigsen,

T. de Jong, A. Lazonder, & S. Barnes (Eds.), *Technology-enhanced learning: Principles and products* (pp. 3–19). Netherlands: Springer.

Ebner, M., Kickmeier-Rust, M. & Holzinger, A. (2008). Utilizing Wiki-Systems in higher education classes: a chance for universal access? *Springer Universal Access in the Information Society, 7*, 199-207.

Evans, C. (2013). Making Sense of Assessment Feedback in Higher Education. *Review of Educational Research, 83,* 70–120.

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist*, *48,* 56–66.

Gibbs, G. & Simpson, C. (2004) Conditions under which assessment supports students' learning, *Learning and Teaching in Higher Education, 1*, 3–31.

Gielen, M., & De Wever, B. (2012). Peer Assessment in a Wiki: Product Improvement, Students' Learning And Perception Regarding Peer Feedback. *Procedia - Social and Behavioral Sciences, 69*, 585-594.

Gielen, M., & De Wever, B. (2015). Structuring the peer assessment process: a multilevel approach for the impact on product improvement and peer feedback quality. *Journal of Computer Assisted Learning, 31*, 435-449.

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction, 20*, 304–315.

Hattie, J., & Gan, M. (2011). Instruction based on feedback. In R. E. Mayer & P. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 249–271). New York: Routledge and Taylor & Francis Group.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 81–112.

Hovardas, T., Tsivitanidou, O. E., & Zacharia, Z. C. (2014). Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. *Computers & Education*, *71*, 133–152.

Kim, M. (2009). The impact of an elaborated assessee's role in peer assessment. *Assessment and Evaluation in Higher Education, 34,* 105–114.

King, A. (2002). Structuring peer interaction to promote high-level cognitive processing. *Theory into Practice*, *41*, 33–39.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*, 254–284.

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts – a conceptual analysis. *Educational Psychology Review*, *18*, 159–185.

Kollar, I., & Fischer, F. (2010). Peer assessment as collabo- rative learning: A cognitive perspective. *Learning and Instruction*, *20*, 344–348.

Kollar, I., Ufer, S., Reichersdorfer, E., Vogel, F., Fischer, F., & Reiss, K. (2014). Effects of collaboration scripts and heuristic worked examples on the acquisition of mathematical argumentation skills of teacher students with different levels of prior achievement. *Learning and Instruction, 32,* 22-36.

Nelson, M. M., & Schunn, C. D. (2008). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science*, *37*, 375–401.

Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment and Evaluation in Higher Education, 35*, 501–517.

Nicol, D., & MacFarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*, 199–218.

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review per- spective. *Assessment & Evaluation in Higher Education*, *39*, 102–122.

Noroozi, O., Weinberger, A., Biemans, H. J. A., Mulder, M., & Chizari, M. (2013). Facilitating argumentative knowledge construction through a transactive discussion script in CSCL. *Computers and Education, 61*, 59-76.

Poverjuc, O., Brooks, V., & Wray, D. (2012). Using peer feedback in a Master's programme: a multiple case study. *Teaching in Higher Education, 17,* 465-477.

Prins, F., Sluijsmans, D., & Kirschner, P. A. (2006). Feedback for general practitioners in training: quality, styles, and preferences. *Advances in Health Sciences Education, 11*, 289-303.

Sadler, D. R. (1998) Formative assessment: revisiting the territory, *Assessment in Education, 5*, 77–84.

Shute, V. J., Hansen, E. G., & Almond, R. G. (2007). An assessment for learning system called ACED: Designing for learning effectiveness and accessibility. (RR-07-26). Princeton, NJ: Educational Testing Service.

Slavin, R. E. (1989). Research on cooperative learning: an international perspective. *Scandinavian Journal of Educational Research, 33*, 231-243.

Sluijsmans, D., Brand-Gruwel, S., van Merriënboer, J. J. G. & Bastiaens, T. (2002). The training of peer assessment skills to promote the development of reflection skills in teacher education, *Studies in Educational Evaluation, 29*, 23–42.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

Strijbos, J. W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in aca- demic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction, 20*, 291–303.

Strijbos, J. W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction, 20*, 265–269.

Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*, 249–276.

Topping, K. J. (2009). Peer assessment. *Theory into Practice, 48*, 20–27.

Tsivitanidou, O. E., Zacharia, Z. C., & Hovardas, T. (2011). Investigating secondary school students' unmediated peer assessment skills. *Learning and Instruction*, 21, 506-519.

Van Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist, 38*, 5–13.

Van Lehn, K.A., Chi, M. T., Baggett, W. , & Murray, R. C. (1995). Progress report: Towards a theory of learning during tutoring. Pittsburgh, PA: Learning Research and Development Center, University of Pittsburg.

Van Steendam, E., Rijlaarsdam, G., Sercu, L., & Van den Berg, H. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher- order peer feedback in EFL. *Learning and Instruction*, *20*, 316–327.

Webb, N. M. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education, 22*, 366-389.

Wecker, C., & Fischer, F. (2011). From guided to self-regulated performance of domain-general skills: the role of peer monitoring during the fading of instructional scripts. Learning and Instruction, 21, 746-756.

Weinberger, A., Stegmann, K., Fischer, F., & Mandl, H. (2007). Scripting argumentative knowledge construction in computer-supported learning environments. In F. Fischer, H. Mandl, J. Haake, & I. Kollar (Eds.), *Scripting computer-supported communication of knowledge—cognitive, computational and educational perspectives* (pp. 191–211). New York: Springer.